PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 : C12Q 1/68

A1

(11) International Publication Number: WO 95/30774

(43) International Publication Date: 16 November 1995 (16.11.95)

(21) International Application Number: PCT/US95/04899

(22) International Filing Date: 24 April 1995 (24.04.95)

(30) Priority Data:
08/239,475 5 May 1994 (05.05.94) US

(71) Applicant: BECKMAN INSTRUMENTS, INC. [US/US]; 2500 Harbor Boulevard, P.O. Box 3100, Fullerton, CA 92634-3100 (US).

(71)(72) Applicant and Inventor: CASKEY, Charles, Thomas [US/US]; 6402 Belmont, Houston, TX 77005 (US).

(72) Inventors: MATSON, Robert, S.; 8324 Candleberry Circle, Orange, CA 92669 (US). COASSIN, Peter, J.; 27626 Paseo Sonata, San Juan Capistrano, CA 92675 (US). RAMPAL, Jang, B.; 20478 Via Canarias, Yorba Linda, CA 92687 (US).

(74) Agent: SPITALS, John, P.; Robbins, Berliner & Carson, 5th floor, 201 N. Figueroa Street, Los Angeles, CA 90012 (US).

(81) Designated States: AU, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published

With international search report.
Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.

(54) Title: OLIGONUCLEOTIDE REPEAT ARRAYS

(57) Abstract

A solid support based hybridization assay is provided which allows for the systematic and reproducible analysis of repeat and tandem repeat oligonucleotide sequences of DNA and RNA by hybridization to a reverse dot blot array comprising strings of such repeats complementary to those found in particular nucleic acid targets (e.g., analyte PCR product). An addressable library (i.e., an indexed set) of complementary repeats is synthesized on a suitable support. Preferably, the support comprises a low fluorescent background support, thereby facilitating the use of non-radioisotopic modes of detection (such as fluorescence of chemiluminescence); particularly suitable in this regard is an aminated polypropylene support or similar material. Preferred arrays permit screening of DNA and RNA samples for complete sets of particular types of nucleotide repeat sequences (e.g., all nucleotide doublet or triplet repeats).

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	GB	United Kingdom	MR	Mauritania
AU	Australia	GE	Georgia	MW	Malawi
BB	Barbados	GN	Guinea	NE	Niger
BE	Belgium	GR	Greece	NL	Netherlands
BF	Burkina Faso	HU	Hungary	NO	Norway
BG	Bulgaria	IE	Ireland	NZ	New Zealand
BJ	Benin	IT	Italy	PL	Poland
BR	Brazil	JP	Japan	PT	Portugal
BY	Belarus	KE	Kenya	RO	Romania
CA	Canada	KG	Kyrgystan	RU	Russian Federation
CF	Central African Republic	KP	Democratic People's Republic of Korea	SD	Sudan
CG	Congo			SE	Sweden
CH	Switzerland	KR	Republic of Korea	SI	Slovenia
CI	Côte d'Ivoire	KZ	Kazakhstan	SK	Slovakia
CM	Cameroon	LI	Liechtenstein	SN	Senegal
CN	China	LK	Sri Lanka	TD	Chad
CS	Czechoslovakia	LU	Luxembourg	TG	Togo
CZ	Czech Republic	LV	Latvia	TJ	Tajikistan
DE	Germany	MC	Monaco	TT	Trinidad and Tobago
DK	Denmark	MD	Republic of Moldova	UA	Ukraine
ES	Spain	MG	Madagascar	US	United States of America
FI	Finland	ML	Mali	UZ	Uzbekistan
FR	France	MN	Mongolia	VN	Viet Nam
GA	Gabon				

OLIGONUCLEOTIDE REPEAT ARRAYS

Background of the Invention

The present invention relates generally to the fields of biochemistry and medicine. In particular, the invention is directed to materials and methods useful in the diagnosis of genetic mutations of clinical relevance.

Short tandem repeats (STR) have been identified in a number of genes. It has been proposed that particular unstable triplet repeat oligonucleotides are correlated with a number of genetic diseases in humans, including Kennedy's disease [La Spada, A. et al., *Nature*, **352**, 77-79 (1991)], fragile-X syndrome [Verkerk, A.J.M.H. et al., *Cell* **65**, 905-914 (1991)], myotonic dystrophy [Fu, Y.H. et al, *Science* **255**, 1256-1258 (1992)], Huntington disease [The Huntington's Disease Collaborative Research Group, *Cell* **72**, 971-983 (1993)] and spinocerebellar ataxia type 1 [Orr, H.T. et al., *Nature Genet.* **4**, 221-226 (1993)]. Similarly, doublet repeats have also been reported to be associated with particular disease states; for example, correlations have been proposed with cystic fibrosis [Chu, C.-S. et al., *Nature Genetics* **3**, 151-156 (1993)] and colorectal cancer [Thibodeau, S.N. et al., *Science* **260**, 816-819 (1993)]. Higher-order repeats, such as tetramers [*see, e.g.*, Gen, M.W. et al., *Genomics* **17**, 770-772 (1993)], have also been identified.

One gene which has been subject of intense scrutiny is the Huntington's disease gene. The trinucleotide hybridization approach was recently utilized to map out tandem repeats across a section of the gene. In this section, 51 triplet repeats spanning a 1.86 Mbp DNA segment were identified by Southern transfer of restriction enzyme digests of a specific cosmid and probing with ^{32}P-labelled oligonucleotide probes [Hummerich, et al., "Distribution of trinucleotide repeat sequences across a 2 Mbp region containing the

Huntington's disease gene," *Human Molecular Genetics* **3**, 73 (1994)].

DNA polymorphisms which arise from allelic differences in the number of repeats have been identified by such terminology as short tandem repeats (STR), variable number of tandem repeats (VNTR), minisatellites (tandem repeats of a short sequence, originally defined as 9-60 bp) and microsatellites (originally defined as 1-5 bp) [McBride, L.J. & O'Neill, M.D., *American Laboratory*, pp. 52-54 (November 1991)]; minisatellites and microsatellites would be considered subclasses of the VNTR. It is estimated that there are up to 500,000 microsatellite repeats distributed throughout the human genome, at an average spacing of 7000 bp. Therefore, it is apparent that most genes will contain VNTR regions and that these regions can be used as genetic markers. For example, VNTRs are currently being used as markers in studies concerned with the inheritance of certain mutations leading to various forms of cancer. Recently, it has been discovered that certain triplet repeat expansions are associated with a predisposition towards certain diseases; a large expansion is typically associated with the onset of the disease. For example, the (CGG) triplet repeat region associated with Fragile X occurs at a frequency of 10-50 repeat units in the normal population, while in those afflicted with the disease the expansion is between 200-2000 repeats.

As it becomes possible to determine whether a particular genotype comprises an unstable repeat and/or is associated with a particular disease state, there is a considerable incentive to develop useful methods to characterize STRs. The heretofore available methods for initial scanning for STRs have generally required time-consuming sequential oligonucleotide hybridizations to filter-bound target DNAs to identify specific STRs [*see*, *e.g.*, Litt, M. and Luty, J.A., *Am. J. Hum. Genet.* **44**, 397-401 (1989); Weber, J.L. and May, P.E., *Am. J. Hum.

Genet. **44**, 388-396 (1989); Fu et al., *supra*]. In particular, the analysis of oligonucleotide repeats is typically carried out at the present time by Southern blotting of restriction fragments followed by hybridization analysis using a specified repetitive sequence probe. Alternatively, it is possible to probe dot blots of the target DNA [Iizuka, et al., *GATA* **10**:2-5 (1993)].

Both of these heretofore-known techniques are time-consuming and tedious for large sample populations. Moreover, multiple probings may be required to identify which repeat might be present. Further, it is often difficult to reproducibly spot or transfer equivalent amounts of DNA to these supports; thus, conventional dot blots and transfers show variation in signal intensity from batch to batch. In addition, any regions of DNA that might become cross-linked to the support (e.g., through UV light) would be inaccessible to probes.

It would be highly useful for clinical investigators to be able to screen large sample populations of patients DNAs in an effective manner. As additional STRs are identified and associated with particular conditions, the need for simple and effective screening methods becomes greater.

PCT published application No. WO 89/10977 describes methods and apparatus for analyzing polynucleotide sequences in which an array of the whole or a chosen part of a complete set of oligonucleotides are bound to a solid support. The different oligonucleotides occupy separate cells of the array and are capable of taking part in hybridization reactions. For studying differences between polynucleotide sequences, the array may comprise the whole or a chosen part of a complete set of oligonucleotides comprising the polynucleotide sequences. While it is suggested that a small array may be useful for many applications, such as the analysis of a gene for mutations, there is no teaching or suggestion

of a specific array or method for using same which would permit the rapid and accurate screening of a wide range of biological materials for tandem repeats. Moreover, the arrays described in WO 89/10977 are designed specifically for use in sequencing by hybridization; the presence of long tandem nucleotide repeats can present a significant problem in attempts to sequence a sample using the methods described in WO 89/10977.

It is an object of the present invention to provide methods and apparatus for rapid and accurate identification of nucleotide tandem repeats in DNA and RNA sequences from a wide variety of sources.

<u>Summary of the Invention</u>

In accordance with the present invention, a solid support based hybridization assay is provided which allows for the systematic and reproducible analysis of repeat and tandem repeat oligonucleotide sequences of DNA and RNA by hybridization to a reverse dot blot array comprising strings of such repeats complementary to those found in particular nucleic acid targets (e.g., analyte PCR product). An addressable library (i.e., an indexed set) of complementary repeats is synthesized on a suitable support. Preferably, the support comprises a low fluorescent background support, thereby facilitating the use of non-radioisotopic modes of detection (such as fluorescence or chemiluminescence); particularly suitable in this regard is an aminated polypropylene support or similar material. Pursuant to a preferred embodiment of the invention, arrays are provided which permit screening of DNA and RNA samples for complete sets of particular types of nucleotide repeat sequences (e.g., all nucleotide doublet or triplet repeats).

Brief Description of the Drawings

The invention may be better understood with reference to the accompanying drawings, in which:

Fig. 1 illustrates a pattern for synthesis of leader sequences and tandem repeat patterns;

Figs. 2(A) - 2(C) illustrate hybridization of arrayed dinucleotide and trinucleotide oligonucleotide repeats using ^{32}P labelled DNA of cosmid 22.3 as a probe, in which Fig. 2(A) represents hybridization stringency at 40°C, Fig. 2(B), hybridization stringency at 50°C , and Fig. 2(C), hybridization stringency at 60°C; and

Fig. 3 illustrates the type and position of STRs indicated by the STR-Strips in 34,977 bp of cosmid 22.3.

Detailed Description of the Invention

In accordance with the present invention, defined repeat and tandem repeat arrays for use in screening nucleic acid targets for the presence of genetic markers generally known as variable number [of] tandem repeats (VNTRs) are synthesized on a suitable support. After hybridization of the target materials with the array, the identity of any tandem repeat sequence(s) in the target materials may be readily ascertained by observing the location(s) at which binding occurs. Pursuant to the present invention, probes are reproducibly synthesized on the surface, freely accessible to target DNA. Moreover, all hybridizations can be rapidly identified under a limited number of stringency conditions.

The arrays of the present invention comprise a predetermined set of oligonucleotides attached to the surface of the solid support. One particularly useful class of tandem repeats for arrays in accordance with the present invention comprises the complete class of 60 tandem triplet repeats (i.e., all possible triplet combinations minus the four homopolymer combinations).

Another useful class of tandem repeats is the complete class of 6 tandem doublet repeats (i.e., the 10 possible doublet combinations of the nucleic acids A, C, G and T minus the four homopolymer combinations). Of course, those skilled in the art would readily appreciate that a wide range of different combinations of repeat elements could also be employed in accordance with the present invention. For example, repeats of a higher order (i.e., repeats of four or more nucleotides) may be useful in some instances. In addition, particular subclasses of any complete class of all possible tandem repeats of a given size may be suitable for carrying out particular types of screenings For purposes of the present invention, all predetermined sets of tandem repeats are contemplated as within the scope of the present invention.

The sequences forming the array may be directly linked to the support. In other embodiments of the arrays of the present invention, the repeat units may be attached to the support by non-repetitive sequences of oligonucleotides or other molecules serving as spacers or linkers to the solid support. In preferred examples of this embodiment, specific leader sequences are encoded on either side of the tandem repeat region in an array format. Depending upon the relative position of the leader sequence a PCR or sequencing primer may be designed. Such primers may then be used to aid in the characterization of the length of the tandem repeat and/or the specific flanking sequences, respectively. In general, a triplet tandem repeat sequence of sufficient length effectively defines two additional tandem repeat sequences; for example, a 21mer complementary to $(ACG)_n$ also hybridizes to $(GAC)_n$ and $(CGA)_n$. By systematically including a degenerate set of leader sequences while reducing the size of the tandem repeat region, hybridization stringency is increased to allow for identification of the combination of leader plus the

tandem repeat; in the example, selectivity of CCC ACG ACG ACG [SEQ ID NO:1] would be observed over, e.g., CCC GAC GAC GAC [SEQ ID NO:2]. Fig. 1 illustrates a set of instructions for synthesis of suitable leader sequences for triplet tandem repeats. Such leader arrays are particularly advantageous for the purpose of identifying leader sequences for use as PCR primers to tandem repeat regions.

The method of the present invention is generally applicable to a wide range of tandem repeat patterns, including higher order tandem repeats. As by definition a tandem repeat consists of at least 2 units of a given oligomer (for example, a dimer or 2mer), then a $(2mer)_n$ wherein n = 2, 3, 4, ... would represent a dinucleotide repeat forming a 4mer, 6mer, 8mer, etc. (e.g., ACAC, ACACAC, ACACACAC, etc.). Similarly, a triplet repeat would be defined as a $(3mer)_n$ and a tetramer repeat as $(4mer)_n$, wherein n represents the number of repeats present. Contemplated as within the scope of the present invention are all tandem repeats of the general formula $(Nmer)_n$ wherein N is an integer greater than 1 representing the number of nucleotides in the repeat pattern and n is an integer representing the number of times the pattern is repeated; in general, the product of N and n is in the range of 4 to about 100, and preferably 6 to about 60.

Higher order tandem repeat combinations representing combinations of two or more individual tandem repeats are also contemplated as within the scope of the present invention; for example, such higher order tandem repeat combinations may include two dimer patterns, a dimer and a triplet, two triplets, etc. In general terms, such repeat combinations may be described as

$$(Nmer)_n(Mmer)_m$$

in which N and M are independently selected integers greater than 1 and represent the number of nucleotides in the respective repeat pattern and n and m are

independently selected integers and represent the number of times the respective pattern is repeated. In the case of three tandem patterns, the structure may be represented as

$$(Nmer)_n(Mmer)_m(Pmer)_p$$

in which P is defined in the same manner as N and M, and p in the same manner as n and m. Moreover, these higher order tandem repeat combinations may also be found in a repeat pattern; such a complex higher order tandem repeat combination may be described as

$$[(Nmer)_n(Mmer)_m)]_x$$

or

$$[(Nmer)_n(Mmer)_m(Pmer)_p]_x$$

in which N, M, P, n, m and p are as previously defined and x is an integer which represents the number of times the $[(Nmer)_n(Mmer)_m)]$ or $[(Nmer)_n(Mmer)_m(Pmer)_p]$ pattern is repeated. Contemplated as within the scope of the present invention are those combinations wherein x(Nn + Mm) or x(Nn + Mm + Pp) is in the range of 4 to about 100, preferably in the range of 6 to about 60. In the simplest case comprising two repeat patterns and wherein x is 1, N and M are both 2 and n and m are both 1; 1 x 2 + 1 x 2 = 4. Table 1 illustrates the construction of higher order repeats in which n and m are both 2.

TABLE 1

Tandem Repeat	$(2mer)_n$	$(3mer)_n$	$(4mer)_n$	$(5mer)_n$	$(6mer)_n$	$(7mer)_n$
$(2mer)_m$	8mer	10mer	12mer	14mer	16mer	18mer
$(3mer)_m$	10mer	12mer	14mer	16mer	18mer	20mer
$(4mer)_m$	12mer	14mer	16mer	18mer	20mer	22mer
$(5mer)_m$	14mer	16mer	18mer	20mer	22mer	24mer
$(6mer)_m$	16mer	18mer	20mer	22mer	24mer	26mer
$(7mer)_m$	18mer	20mer	22mer	24mer	26mer	28mer

Thus, for example, the 8mer comprising a $(2mer)_n = (AC)_2$ and a $(2mer)_m = (AT)_2$ would have the formula ACACATAT;

similarly, the 12mer comprising a $(3mer)_n = (ACG)_2$ and a $(3mer)_m = (ATC)_2$ would have the formula ACGACGATCATC [SEQ ID NO:3]. A complex higher order repeat $[(AC)_n(AT)_m]_x$ in which n and m are 1 and x is 2 would have the formula ACATACAT; where x is 3, the formula would be ACATACATACAT [SEQ ID NO:4]. Those skilled in the art would readily appreciate the variety of simple and higher-order tandem repeat patterns falling within the scope of the present invention.

The methods and apparatus in accordance with the present invention take advantage of the fact that under appropriate conditions oligonucleotides form base paired duplexes with oligonucleotides which have a complementary base sequence. The stability of the duplex is dependent on a number of factors, including the length of the oligonucleotides, the base composition, and the composition of the solution in which hybridization is effected. The effects of base composition on duplex stability may be reduced by carrying out the hybridization in particular solutions, for example in the presence of high concentrations of tertiary or quaternary amines.

The thermal stability of the duplex is also dependent on the degree of sequence similarity between the sequences. By carrying out the hybridization at temperatures close to the anticipated Tm's of the type of duplexes expected to be formed between the target material(s) and the oligonucleotides bound to the array, the rate of formation of mismatched duplexes may be substantially reduced.

The number of repeats in the tandem sequences attached to the array may vary over a broad range from the minimum of two necessary to constitute a repeat to a maximum on the order of about 50. Of course, the optimum range for the number of tandem repeats in any given instance is dependent upon a number of factors, including in particular the composition and the length of the

repeat. In general, the Tm of the complex formed between a given sequence in the target material and the complementary sequence in the array increases as the length of the sequences increase; however, only minor increases in Tm are observed once the sequences have reached a length of about 50-60 bases. The sequences on the array contain at least four bases (the minimum for a repeat of a doublet pattern). It is generally preferred that the sequences on the array comprise at least about 6 bases, more preferably at least about 10 bases, and most preferably on the order of about 15 to about 60 bases. As previously noted, there is little practical advantage in using sequences substantially longer than about 60 bases; nonetheless, extended sequences of up to about 100 bases in length (corresponding to, e.g., 50 repeats of a doublet sequence) and longer are also contemplated as within the scope of the present invention.

In addition, in accordance with preferred embodiments of the invention the length of each sequence employed in the array may be selected to as to optimize binding of target materials to the array. For any given tandem repeat sequence, an optimum length for use with any particular target material under specified screening conditions may be determined empirically. Thus, the length for each individual element of the set of tandem repeats comprising the array may be optimized for the screening of particular target materials under specific conditions (e.g., at a given temperature).

A wide variety of array formats may be employed in accordance with the present invention. One particularly useful format is a linear array of oligonucleotide bands, generally referred to in the art as a dipstick. Another suitable format comprises a two-dimensional pattern of discrete cells (e.g., 4096 squares in a 64 by 64 array). Of course, as would be readily appreciated by those skilled in the art, other array formats (e.g., circular)

would be equally suitable for use in accordance with the present invention. While arrays may be prepared on a variety of materials including glass plates, it is presently preferred to use an organic polymer medium. As used herein, the term "organic polymer" is intended to mean a support material which is most preferably chemically inert under conditions appropriate for biopolymer synthesis and which comprises a backbone comprising various elemental substituents including, but not limited to, hydrogen, carbon, oxygen, fluorine, chlorine, bromine, sulfur and nitrogen. Representative polymers include, but are not limited to, the following: polypropylene, polyethylene, polybutylene, polyisobutylene, polybutadiene, polyisoprene, polyvinylpyrrolidone, polytetrafluoroethylene, polyvinylidene difluoride, polyfluoroethylene-propylene, polyethylene-vinyl alcohol, polymethylpentene, polychlorotrifluoroethylene, polysulfones, and blends and copolymers thereof. As used herein, the term "medium" is intended to mean the physical structural shape of the polymer. Thus, medium can be generally defined as polymer films (i.e., polymers having a substantially non-porous surface); polymer membranes (i.e., polymers having a porous surface); polymer filaments (e.g., mesh and fabrics); polymer beads; polymer foams; polymer frits; and polymer threads. Preferably, the polymer medium is a thread, membrane or film; most preferably, the polymer medium is a film. An exemplary organic polymer medium is a polypropylene sheet having a thickness on the order of about 1 mil (0.001 inch), although the thickness of the film is not critical and may be varied over a fairly broad range. Particularly preferred for preparation of arrays at this time are biaxially oriented polypropylene (BOPP) films; in addition to their durability, BOPP films exhibit a low background fluorescence.

The array formats of the present invention may be included in a variety of different types of device. As

used herein, the term "device" is intended to mean any device to which the solid support can be affixed, such as microtiter plates, test tubes, inorganic sheets, dipsticks, etc. For example, when the solid support is a polypropylene thread, one or more polypropylene threads can be affixed to a plastic dipstick-type device; polypropylene membranes can be affixed to glass slides. The particular device is, in and of itself, unimportant. All that is necessary is that the solid support can be affixed thereto without affecting the functional behavior of the solid support or any biopolymer adsorbed thereon, and that the device is stable to any materials into which the device is introduced (e.g., clinical samples, etc.).

The arrays of the present invention may be prepared by a variety of approaches which are known to those working in the field. Pursuant to one type of approach, the complete sequences are synthesized separately and then attached to a solid support. However, it is presently considered particularly advantageous to synthesize the sequences directly onto the support to provide the desired array. Suitable methods for covalently coupling oligonucleotides to a solid support and for directly synthesizing the oligonucleotides onto the support would be readily apparent to those working in the field; a summary of suitable methods may be found in, e.g., Matson, R.S. et al., *Analytical Biochem.* **217**, 306-310 (1994), hereby incorporated by reference. Advantageously, the oligonucleotides are synthesized onto the support using conventional chemical techniques as heretofore employed for preparing oligonucleotides on solid supports comprising, e.g., controlled pore size glass (CPG), as described for example in PCT applications WO 85/01051 and WO 89/10977, or polypropylene, as described in copending U.S. patent application Serial No. 07/091,100, which has been assigned to the assignee of the present application and is herein incorporated by reference. Pursuant to one preferred approach, a

polypropylene support (for example, a biaxially-oriented polypropylene) is first surface aminated by exposure to an ammonia plasma generated by radiofrequency plasma discharge. The reaction of a phosphoramidite-activated nucleotide with the aminated membrane followed by oxidation with, e.g., iodine provides a base stable amidate bond to the support.

As described in U.S. patent application Serial No. 08/144,954 filed October 28, 1993, which has been commonly assigned to the assignee of the present invention and is incorporated by reference herein, a suitable array may advantageously be produced using automated means to synthesize oligonucleotides in the cells of the array by laying down the precursors for the four bases in a predetermined pattern. Briefly, a multiple-channel automated chemical delivery system is employed to create oligonucleotide probe populations in parallel rows (corresponding in number to the number of channels in the delivery system) across the substrate. Following completion of oligonucleotide synthesis in a first (1°) direction, the substrate may then be rotated by 90° to permit synthesis to proceed within a second (2°) set of rows that are now perpendicular to the first set. This process creates a multiple-channel array whose intersection generates a plurality of discrete cells. Table 2 describes an exemplary vertical array of 64 oligonucleotides consisting of 60 triplet tandem repeat sequences (21mers) and dinucleotide tandem repeat sequences (20mers).

For the example of the preferred array comprising specific leader sequences as described in Fig. 1 (and as more fully described in the above-noted U.S. patent application Serial No. 08/144,954), all of the degenerate triplet repeats (n=64, including homopolymers) are synthesized in a first direction (1° synthesis) in the 64 channels in Cycles 1-3 as 3′ Leader Sequences (LLL). For example, lane 1 is AAA, lane 2 AAC, lane 5 ACA and lane

64 TTT. Then, the film is rotated 90° to perform synthesis in the second direction (2° synthesis or cross-synthesis) in Cycles 4-15. All 64 triplet tandem repeat sequences of $(NNN)_4$ are then produced. The 2-dimensional array created thereby is the product of a bidirectional synthesis and comprises 4096 discrete cells containing 15mer oligonucleotide products $(LLL)(NNN)_4$, in which the Leader Sequence is placed at the 3'-end of each completed oligonucleotide. Thus, the following sequences would be found at the indicated cell positions:

Cell 1,1'	$(AAA)(AAA)_4$	[SEQ ID NO:5]
Cell 1,2'	$(AAA)(AAC)_4$	[SEQ ID NO:6]
Cell 5,1'	$(ACA)(AAA)_4$	[SEQ ID NO:7]
Cell 64,1'	$(TTT)(AAA)_4$	[SEQ ID NO:8]
Cell 64,64'	$(TTT)(TTT)_4$	[SEQ ID NO:9]

This type of array comprising leader sequences (at either the 5' or 3' end) is particularly preferred in accordance with the present invention.

In order to accommodate a suitably large array, the pixel size should be as small as possible. Cells having a width on the order of about 10 μm to about 1 mm would be particularly suitable. In one preferred embodiment of the invention, arrays with a pixel width of about 500 μm are prepared on biaxially-oriented polypropylene.

Pursuant to the present invention, there are also provided methods for screening DNA and RNA samples comprising labelling the samples to form labelled material, applying the labelled material under suitable hybridization conditions to an array as described herein, and observing the location of the label on the surface associated with particular members of the set of oligonucleotides. Identification of the cell(s) in which binding occurs permits a rapid and accurate identification of any nucleotide repeats present in the sample from which the probes are derived.

In a hybridization reaction in accordance with the present invention, the array is explored by the labelled

target material in essentially the same manner as a labelled probe is employed to screen, e.g., a DNA library containing a gene complementary to the probe. The target material may suitably comprise labelled sequences amplified from genomic DNA by the polymerase chain reaction (PCR), a mRNA population, or a partial or complete set of oligonucleotides from one or more chromosomes or an entire genome. To prepare the target material, the sample may be degraded to form fragments; where appropriate, the degraded material may then be sorted (for example, by electrophoresis on a sequencing gel) to provide a set of oligomers having a specific length.

The target material is then labelled to facilitate detection of duplex formation. Suitably, conventional methods for end-labelling of oligomers are employed. Both radioactive and fluorescent labelling methods would be suitable for use in accordance with the present invention. Commonly-employed techniques routinely permit the introduction of label into a significant fraction of the target materials. Using conventional methods for labelling oligomers with ^{32}P, for example, the radioactive yield of any individual oligomer even from a total human genome could be more than 10^4 dpm/mg of total DNA. For detection, only a small fraction of the labelled material would be necessary for hybridization to a pixel of a size within the preferred range specified herein. Hybridization conditions for a given combination of array and target material can routinely be optimized in an empirical manner to be close to the Tm of the expected duplexes, thereby maximizing the discriminating power of the method. Autoradiography (in particular, with ^{32}P) may cause image degradation which may be a limiting factor determining resolution; the limit for silver halide films is about 25 μm. Accordingly, the use of fluorescent probes (in particular, in conjunction with an array prepared on a low-fluorescence solid support) is

presently preferred. In view of the low background fluorescence of the preferred biaxially-oriented polypropylene substrate, fluorescence-based labelling techniques may advantageously be employed with arrays on such a substrate. With either type of labeled target material, the substantial excess in bound oligonucleotides of the array makes it possible to operate at conditions close to equilibrium with most types of target materials contemplated herein.

As would be readily understood by those skilled in the art, the chosen conditions of hybridization must be such as to permit discrimination between exactly matched and mismatched oligonucleotides. Hybridization conditions may be initially chosen to correspond to those known to be suitable in standard procedures for hybridization to filters and then optimized for use with the arrays of the present invention; moreover, conditions suitable for hybridization of one type of target material would appropriately be adjusted for use with other target materials for the same array. In particular, it is appropriate to control temperature closely (preferably, to better than about ±1° C) to substantially eliminate formation of duplexes between sequences other than identical sequences. Particularly when the length of the oligonucleotide in the target materials is small, it is necessary to be able to distinguish between slight differences in the rate and/or extent of hybridization.

A variety of heretofore known hybridization solvents may suitably be employed, the choice of solvent for particular hybridizations being dependent on a number of considerations. For example, G:C base pairs are more stable than A:T base pairs in 1 M NaCl; thus, the Tm of double-stranded oligonucleotides with a high G + C content will be higher than corresponding oligonucleotides with a high A + T content. These effects are of course particularly pronounced in sequences comprising tandem nucleotide repeats. In order

to compensate for this discrepancy, a variety of approaches may be employed. For example, the amount of oligonucleotide applied to the surface of the support may be varied in dependence on the nucleotide composition of the bound oligomer. Further, computer means employed to analyze data from hybridization experiments may be programmed to make compensations for variations in nucleotide compositions. Another expedient (which may be employed instead of or in addition to those already mentioned) is the use of a chaotropic hybridization solvent, such as a ternary or quaternary amine. In this regard, tetramethylammonium chloride (TMACl) at concentrations in the range of about 2 M to about 5.5 M is particularly suitable; at TMACl concentrations around 3.5 to 4 M, the Tm dependence on nucleotide composition is substantially reduced. In addition, the choice of hybridization salt has a major effect on overall hybridization yield; for example, TMACl at concentrations up to 5 M can increase the overall hybridization yield by a factor of up to 30 or more (depending to some extent on the nucleotide composition) compared to 1 M NaCl. Finally, as previously noted, the length of the oligonucleotides attached to the array may be varied so as to optimize hybridization under the particular conditions employed. As previously noted, it would be a routine matter for those working in the field to optimize hybridization conditions for any given combination of target materials and array.

Hybridization is typically carried out with a very large excess of the bound oligonucleotides over what is found in the target. In preferred embodiments of the invention, it is possible in some cases to distinguish between hybridization involving single and multiple occurrences of the target sequence, as yield is proportional to concentration at all stages in the reaction.

In accordance with another embodiment of the present invention, an array as described herein may be employed to selectively isolate and size variable number of tandem repeats (VNTRs). This is accomplished by preparing a sample comprising VNTRs in a manner known per se [*see, e.g.*, McBride & O'Neill, *supra*], capturing the VNTRs on the array, selectively dissociating the hybrid and eluting the VNTR from the support. A selected lane from the array may be cut out of the support, the ssDNA eluted therefrom, the number of copies thereof increased by PCR amplification and size analysis conducted by a conventional technique (e.g., gel electrophoresis against DNA size markers). The presence of large molecular weight strands would indicate an increase in mutational frequency (i.e., higher orders of tandem repeat regions).

The invention may be better understood with reference to the accompanying example that is intended for purposes of illustration only and should not be construed as in any sense limiting the scope of the present invention as defined in the claims appended hereto.

<u>Example</u>

Oligonucleotides were synthesized directly from monomers onto a 6.6 x 6.6 cm sheet of aminated polypropylene substrate using standard CED-phosphoramidite chemistries. A specially designed 64 channel chemical delivery system (Southern Array Maker™, Beckman Instruments) as described in co-pending U.S. patent application Serial No. 08/144,954 was utilized to prepare the discrete oligonucleotide sequences in parallel rows across the polypropylene substrate. Polypropylene film was surface aminated by a radiofrequency discharge into ammonia gas as described in Matson et al., *supra*. The plasma-aminated film was then placed in the synthesizer. Standard phosphoramidite chemistry was performed in each of the 64 channels to create 64 different oligonucleotide sequences on the

film. The substrate was then cut into 0.5 cm widths perpendicular to the oligonucleotide rows to produce a panel of 64 tandem repeat sequences. For the present study 60 trinucleotide (21mers) and 4 dinucleotide tandem (20mers) repeat sequences were arrayed on a vertical order shown in Table 2. The arrayed trinucleotide repeat set represents all triplet frames except $(AAA)_7$ [SEQ ID NO:10], $(CCC)_7$ [SEQ ID NO:11], $(GGG)_7$ [SEQ ID NO:12], and $(TTT)_7$ [SEQ ID NO:13] in 3'-->5' direction as well as minus strand orientation.

In order to confirm that all sequences were fully represented on the panel a series of complementary probes were prepared that would verify each sequence by row position on the strip. As each triplet repeat (n) of a sufficient length in fact reads the n-1 and n-2 frames as well, only 20 probe sequences were required to identify the 60 triplet repeats on the panel. For example, row 2 containing the sequence $(AAC)_7$ [SEQ ID NO:14] also represents $(ACA)_6$ [SEQ ID NO:15] and $(CAA)_6$ [SEQ ID NO:16] (sequences found at row 5 and row 17, respectively); thus, a single tandem repeat probe which has been labeled will identify the sequences at row positions 2, 5 and 17. Four additional probe sequences were required to identify the 4 doublet tandem repeats synthesized on the strip. It is also possible to combine the sequences of non-complementary probes (i.e., those that will not self-hybridize or form hairpin loops) to reduce the total number of probes necessary to read all row positions. The following ^{32}P 5'-end labeled probes were prepared and used to identify all row positions 1-64 (listed in Table 2) on the panel:

$(TGC)_6$;	[SEQ ID NO:17]
$(TGG)_4(TG)_6(TTA)_4$;	[SEQ ID NO:18]
$(TGG)_4(TGA)_4$;	[SEQ ID NO:19]
$(TCG)_6$;	[SEQ ID NO:20]
$(GGC)_5$;	[SEQ ID NO:21]
$(CG)_4(GAC)_4$;	[SEQ ID NO:22]

$(TCA)_4(TAA)_4(TAC)_4$; [SEQ ID NO:23]
$(CCA)_4(CAA)_4$; [SEQ ID NO:24]
$(TTC)_4(TC)_6(TCC)_4$; [SEQ ID NO:25]
$(GAA)_4(GA)_6(GGA)_4$; and [SEQ ID NO:26]
$(GCC)_4(GCA)_4$ [SEQ ID NO:27].

A number of different test DNAs were employed. This included a 5'-$(CAG)_7$-3' oligonucleotide [SEQ ID NO:28] and 200 bp PCR fragments containing a trinucleotide short tandem repeat of $(CTG)_{11}$ [SEQ ID NO:29] [Fu, Y.H., et al, *Science* **255**, 1256-1258 (1992)] generated from human genomic DNA of a wild type individual. An 800 bp PCR fragment containing a $(CAG)_{10}$ [SEQ ID NO:30] repeat and a 3.0 kb PCR fragment containing a tandemly combined $(GCA)_8$+$(GCG)_4$ repeat [SEQ ID NO:31] were generated from cDNA clones G13 and A12, respectively, recently isolated in a new cDNA identification system [Lee, C.C., et al., *Am. J. Hum. Genet.* **53** (Suppl.), 1321 (1993)]. DNA samples of the cosmid MDY2 [Fu, Y.H., et al, *Science* **255**, 1256-1258 (1992)] containing the entire myotonin protein kinase gene and cosmid 22.3 containing the FMR-1 gene [Verkerk, A.J.M.H., et al., *Cell* **65**, 905-914 (1991)] have also been utilized as target materials to evaluate the test strips.

STRs were identified using the GCG sequence analysis software package (Genetics Computer Group, Inc. 1991). 11,613 bp of cosmid MDY2 (GenBank accession L00727) and 61,612 bp of a contiguous sequence containing the complete 34,977 bp sequence of cosmid 22.3 were searched for STRs.

TABLE 2

Vertical array of 64 oligonucleotides consisting of 60 triplet tandem repeat sequences (21mers) and dinucleotide tandem repeat sequences (20mers) on a polypropylene substrate

#	Oligonucleotides 3' - - - > 5'	Seq ID No.	#	Oligonucleotides 3' - - - > 5'	Seq ID No.
1	AC AC AC AC AC AC AC AC AC AC	32	33	GAA GAA GAA GAA GAA GAA GAA	62
2	AAC AAC AAC AAC AAC AAC AAC	14	34	GAC GAC GAC GAC GAC GAC GAC	63
3	AAG AAG AAG AAG AAG AAG AAG	33	35	GAG GAG GAG GAG GAG GAG GAG	64
4	AAT AAT AAT AAT AAT AAT AAT	34	36	GAT GAT GAT GAT GAT GAT GAT	65
5	ACA ACA ACA ACA ACA ACA ACA	35	37	GCA GCA GCA GCA GCA GCA GCA	66
6	ACC ACC ACC ACC ACC ACC ACC	36	38	GCC GCC GCC GCC GCC GCC GCC	67
7	ACG ACG ACG ACG ACG ACG ACG	37	39	GCG GCG GCG GCG GCG GCG GCG	68
8	ACT ACT ACT ACT ACT ACT ACT	38	40	GCT GCT GCT GCT GCT GCT GCT	69
9	AGA AGA AGA AGA AGA AGA AGA	39	41	GGA GGA GGA GGA GGA GGA GGA	70
10	AGC AGC AGC AGC AGC AGC AGC	40	42	GGC GGC GGC GGC GGC GGC GGC	71
11	AGG AGG AGG AGG AGG AGG AGG	41	43	AG AG AG AG AG AG AG AG AG AG	72
12	AGT AGT AGT AGT AGT AGT AGT	42	44	GGT GGT GGT GGT GGT GGT GGT	73
13	ATA ATA ATA ATA ATA ATA ATA	43	45	GTA GTA GTA GTA GTA GTA GTA	74
14	ATC ATC ATC ATC ATC ATC ATC	44	46	GTC GTC GTC GTC GTC GTC GTC	75
15	ATG ATG ATG ATG ATG ATG ATG	45	47	GTG GTG GTG GTG GTG GTG GTG	76
16	ATT ATT ATT ATT ATT ATT ATT	46	48	GTT GTT GTT GTT GTT GTT GTT	77
17	CAA CAA CAA CAA CAA CAA CAA	47	49	TAA TAA TAA TAA TAA TAA TAA	78
18	CAC CAC CAC CAC CAC CAC CAC	48	50	TAC TAC TAC TAC TAC TAC TAC	79
19	CAG CAG CAG CAG CAG CAG CAG	28	51	TAG TAG TAG TAG TAG TAG TAG	80
20	CAT CAT CAT CAT CAT CAT CAT	49	52	TAT TAT TAT TAT TAT TAT TAT	81
21	CCA CCA CCA CCA CCA CCA CCA	50	53	TCA TCA TCA TCA TCA TCA TCA	82
22	CG CG CG CG CG CG CG CG CG CG	51	54	TCC TCC TCC TCC TCC TCC TCC	83
23	CCG CCG CCG CCG CCG CCG CCG	52	55	TCG TCG TCG TCG TCG TCG TCG	84
24	CCT CCT CCT CCT CCT CCT CCT	53	56	TCT TCT TCT TCT TCT TCT TCT	85
25	CGA CGA CGA CGA CGA CGA CGA	54	57	TGA TGA TGA TGA TGA TGA TGA	86
26	CGC CGC CGC CGC CGC CGC CGC	55	58	TGC TGC TGC TGC TGC TGC TGC	87
27	CGG CGG CGG CGG CGG CGG CGG	56	59	TGG TGG TGG TGG TGG TGG TGG	88
28	CGT CGT CGT CGT CGT CGT CGT	57	60	TGT TGT TGT TGT TGT TGT TGT	89
29	CTA CTA CTA CTA CTA CTA CTA	58	61	TTA TTA TTA TTA TTA TTA TTA	90
30	CTC CTC CTC CTC CTC CTC CTC	59	62	TTC TTC TTC TTC TTC TTC TTC	91
31	CTG CTG CTG CTG CTG CTG CTG	60	63	TTG TTG TTG TTG TTG TTG TTG	92
32	CTT CTT CTT CTT CTT CTT CTT	61	64	CT CT CT CT CT CT CT CT CT CT	93

Oligonucleotide probes were end-labelled with ^{32}P-Gamma-dATP under standard conditions [Sambrook, J., et al., *Molecular Cloning: A Laboratory Manual*, Second Edition. Cold Spring Harbor Laboratory Press, Cold Spring Harbor (1989)]. Double stranded DNA was radiolabelled with ^{32}P-Alpha-dCTP using a Pharmacia random priming labelling kit according to the manufacturer's instructions. To improve the labelling reaction cosmid DNA was digested with EcoRI prior to radiolabelling. The test strips were hybridized without prehybridization in plastic bags containing 6 x SSCP (saline, sodium citrate-phosphate buffer) and 0.01 % sodium dodecyl sulfate (SDS) for 16 hrs. Only target-specific binding to the polypropylene membranes was observed, eliminating the need for a prehybridization step. The specific activity of the radiolabelled probes was adjusted to 5 x 10^6 cpm/ml hybridization solution. After hybridization the test strips were washed in 2 x SSCP, 0.01 % SDS for 20 min. A variety of hybridization and wash temperatures was employed, as hereinafter described. Autoradiograms were developed after 5 minutes to 6 hours exposure at -70° C. The resulting signals on the test strips were evaluated visually.

As expected, the synthetic $(CAG)_7$ [SEQ ID NO:28] oligonucleotide probe hybridized specifically at 60°C to three rows of the array. These corresponded to the oligonucleotide repeats $(CGT)_7$ [SEQ ID NO:57], $(GTC)_7$ [SEQ ID NO:69], and $(TCG)_7$ [SEQ ID NO:84], respectively.

Using double-stranded DNA of 200 bp and 800 bp containing a $(CTG)_{11}$ [SEQ ID NO:29] or a $(CAG)_{10}$ [SEQ ID NO:30] repeat resulted in a pattern of 6 bands corresponding to $(ACG)_7$ [SEQ ID NO:37], $(CGA)_7$ [SEQ ID NO:54], $(CGT)_7$ [SEQ ID NO:57], $(GAC)_7$ [SEQ ID NO:63], $(GTC)_7$ [SEQ ID NO:75], and $(TCG)_7$ [SEQ ID NO:84] - i.e., the sense and antisense orientations. Differences in the signal intensity were observed between the various triplet-representing lanes.

Using the 3 kb PCR fragment containing a combined repeat $(GCA)_8$+$(GCG)_4$ [SEQ ID NO:31] to probe the test strips resulted in a complete set of six oligonucleotides - $(ACG)_7$ [SEQ ID NO:37], $(CGA)_7$ [SEQ ID NO:54], $(CGT)_7$ [SEQ ID NO:57], $(GAC)_7$ [SEQ ID NO:63], $(GTC)_7$ [SEQ ID NO:75], and $(TCG)_7$ [SEQ ID NO:84]. This set represents the six different frame shifts for the $(GCA)_8$ [SEQ ID NO:94] repeat. Additionally, the signals found with $(CCG)_7$ [SEQ ID NO:52], $(CGC)_7$ [SEQ ID NO:55], and $(GCC)_7$ [SEQ ID NO:67]were evident for the 3'--->5' directed frame of the $(GCG)_4$ [SEQ ID NO:95] repeat. No signal was detected under these conditions for the reversed direction indicated by $(GGC)_7$ [SEQ ID NO:71], $(GCG)_7$ [SEQ ID NO:68], and $(CGC)_7$ [SEQ ID NO:55].

Using cosmid MDY2 with an insert size of 31 kb as a probe, a band pattern was observed indicative of the presence of $(CAG)_n$, $(GCC)_n$, and $(CCT)_n$ repeats, respectively. For the $(CCT)_n$, only one direction of the oligonucleotide frame as represented by $(CCT)_7$ [SEQ ID NO:53], $(CTC)_7$ [SEQ ID NO:59], and $(TCC)_7$ [SEQ ID NO: 83] was found to hybridize. A search of 11,613 nt available sequence information (GenBank accession L00727) of cosmid MDY2 revealed the presence of all types of triplet repeats identified by the test strip (Table 3). The repeated triplet numbers vary from 3 for the CCT and GCG type repeats to 11 for the CTG repeat.

TABLE 3

Position (Nucleotide #)	STR
809-817	$(CCT)_3$
8,172-8,180	$(CCT)_3$
9,093-9,101	$(GGA)_3$
10,364-10,372	$(GCC)_3$
10,677-10,709	$(CTG)_{11}$ [SEQ ID NO:29]

The influence of temperature on the STR detection

was evaluated by hybridizing cosmid 22.3 at 40°C, 50°C, and 60°C to the test strips. The strips were then washed at the temperatures used for the respective hybridizations. At 40°C, a band pattern was obtained indicative of $(CA)_n$, $(ACC)_n$, $(CCT)_n$ and $(GCC)_n$ type repeats, respectively (Fig. 2A); for the triplet repeats, only one set of signals representing one direction of the oligonucleotide frame was observed. The pattern at 50°C was also specific for $(CA)_n$, $(ACC)_n$, and $(GCC)_n$ type repeats (Fig. 2B); however, unlike the pattern at 40°C the signals representing a $(TCC)_n$ type repeat disappeared and additional bands indicative of a $(CGT)_n$ type repeat occurred; again, for the trinucleotide repeats only one set of signals representing one direction of the respective oligonucleotide frame was found. At 60°C only signals representing $(CA)_n$ and $(CCG)_n$ type repeats persisted (Fig. 2C). Under these hybridization conditions a full set of the expected 6 bands evident for a $(CCG)_n$ type repeat was observed.

All types of STRs indicated by the test strips were found to be present in 34,977 bp sequence available from cosmid 22.3 (Fig. 3). They range from a single repeat unit of $(CCT)_3$ and $(AGC)_3$ up to 18 repeat units for a (TG)-dinucleotide repeat. There was no unspecific hybridization signal observed. AT-rich repeats also occurring in the sequence in three or less repeat units were not detected by the test strips under the hybridization conditions used.

The array was designed to represent trinucleotide repeats by all three possible frames in 3'-->5' direction, as well as in the reverse direction. Thus, using single stranded DNA a complementary sequence to a given trinucleotide repeat should result in three signals on the test strips; using double stranded DNA six respective bands for a given repeat should occur. For the four dinucleotide repeats only one frame was used for each type. Using this reverse blotting system, the

obtained band pattern provided qualitatively the precise identification of previously known STRs in DNA samples of various complexities between 21 bp - 34,977 bp. Moreover, there was no random or cross hybridization to unspecific sequences observed. Based on the Tm and size of the STRs as well as possible influences by flanking sequences, varying the hybridization stringency can enhance the specificity.

From the foregoing description, one skilled in the art can readily ascertain the essential characteristics of the invention and, without departing from the spirit and scope thereof, can adapt the invention to various usages and conditions. Changes in form and substitution of equivalents are contemplated as circumstances may suggest or render expedient, and although specific terms have been employed herein, they are intended in a descriptive sense and not for purposes of limitation.

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Beckman Instruments, Inc.

(ii) TITLE OF INVENTION: OLIGONUCLEOTIDE REPEAT ARRAYS

(iii) NUMBER OF SEQUENCES: 95

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Robbins, Berliner & Carson
(B) STREET: 201 North Figueroa Street, Suite 500
(C) CITY: Los Angeles
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 90012

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.25

(vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Spitals, John P.
(B) REGISTRATION NUMBER: 29,215
(C) REFERENCE/DOCKET NUMBER: 5727-110

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (213) 977-11001
(B) TELEFAX: (213) 977-1003

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

CCCACGACGA CG 12

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

CCCGACGACG AC 12

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ACGACGATCA TC 12

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

ACATACATAC AT 12

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AAAAAAAAAA AAAAA 15

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

AAAAACAACA ACAAC 15

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

ACAAAAAAAA AAAAA 15

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

TTTAAAAAAA AAAAA 15

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

TTTTTTTTTT TTTTT 15

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AAAAAAAAAA AAAAAAAAAA A 21

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CCCCCCCCCC CCCCCCCCCC C 21

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGGGGGGGGG GGGGGGGGGG G 21

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TTTTTTTTTT TTTTTTTTTT T 21

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

AACAACAACA ACAACAACAA C 21

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

ACAACAACAA CAACAACA 18

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

CAACAACAAC AACAACAA 18

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

TGCTGCTGCT GCTGCTGC 18

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

TGCTGCTGCT GCTGTGTGTG TGTGTTATTA TTATTA 36

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

TGGTGGTGGT GGTGATGATG ATGA 24

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 18 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TCGTCGTCGT CGTCGTCG 18

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GGCGGCGGCG GCGGC 15

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

CGCGCGCGGA CGACGACGAC 20

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

TCATCATCAT CATAATAATA ATAATACTAC TACTAC 36

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

CCACCACCAC CACAACAACA ACAA 24

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

TTCTTCTTCT TCTCTCTCTC TCTCTCCTCC TCCTCC 36

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

GAAGAAGAAG AAGAGAGAGA GAGAGGAGGA GGAGGA 36

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GCCGCCGCCG CCGCAGCAGC AGCA 24

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CAGCAGCAGC AGCAGCAGCA G 21

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 33 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CTGCTGCTGC TGCTGCTGCT GCTGCTGCTG CTG 33

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CAGCAGCAGC AGCAGCAGCA GCAGCAGCAG 30

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

GCAGCAGCAG CAGCAGCAGC AGCAGCGGCG GCGGCG 36

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

ACACACACAC ACACACACAC 20

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

AAGAAGAAGA AGAAGAAGAA G 21

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

AATAATAATA ATAATAATAA T 21

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

ACAACAACAA CAACAACAAC A 21

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

ACCACCACCA CCACCACCAC C

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

ACGACGACGA CGACGACGAC G 21

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

ACTACTACTA CTACTACTAC T 21

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

AGAAGAAGAA GAAGAAGAAG A 21

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

AGCAGCAGCA GCAGCAGCAG C

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

AGGAGGAGGA GGAGGAGGAG G 21

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

AGTAGTAGTA GTAGTAGTAG T 21

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

ATAATAATAA TAATAATAAT A 21

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

ATCATCATCA TCATCATCAT C

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

ATGATGATGA TGATGATGAT G 21

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATTATTATTA TTATTATTAT T 21

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

CAACAACAAC AACAACAACA A 21

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

CACCACCACC ACCACCACCA C

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

CATCATCATC ATCATCATCA T 21

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

CCACCACCAC CACCACCACC A

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CGCGCGCGCG CGCGCGCGCG 20

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

CCGCCGCCGC CGCCGCCGCC G 21

(2) INFORMATION FOR SEQ ID NO:53:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CCTCCTCCTC CTCCTCCTCC T 21

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

CGACGACGAC GACGACGACG A

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CGCCGCCGCC GCCGCCGCCG C 21

(2) INFORMATION FOR SEQ ID NO:56:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

CGGCGGCGGC GGCGGCGGCG G 21

(2) INFORMATION FOR SEQ ID NO:57:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

CGTCGTCGTC GTCGTCGTCG T 21

(2) INFORMATION FOR SEQ ID NO:58:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

CTACTACTAC TACTACTACT A

(2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

CTCCTCCTCC TCCTCCTCCT C 21

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

CTGCTGCTGC TGCTGCTGCT G 21

(2) INFORMATION FOR SEQ ID NO:61:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CTTCTTCTTC TTCTTCTTCT T 21

(2) INFORMATION FOR SEQ ID NO:62:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

GAAGAAGAAG AAGAAGAAGA A

(2) INFORMATION FOR SEQ ID NO:63:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GACGACGACG ACGACGACGA C 21

(2) INFORMATION FOR SEQ ID NO:64:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

GAGGAGGAGG AGGAGGAGGA G 21

(2) INFORMATION FOR SEQ ID NO:65:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

GATGATGATG ATGATGATGA T 21

(2) INFORMATION FOR SEQ ID NO:66:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GCAGCAGCAG CAGCAGCAGC A

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

GCCGCCGCCG CCGCCGCCGC C 21

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCGGCGGCGG CGGCGGCGGC G

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

GCTGCTGCTG CTGCTGCTGC T

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GGAGGAGGAG GAGGAGGAGG A 21

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

GGCGGCGGCG GCGGCGGCGG C

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

AGAGAGAGAG AGAGAGAGAG 20

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

GGTGGTGGTG GTGGTGGTGG T 21

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

GTAGTAGTAG TAGTAGTAGT A 21

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GTCGTCGTCG TCGTCGTCGT C

(2) INFORMATION FOR SEQ ID NO:76:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

GTGGTGGTGG TGGTGGTGGT G 21

(2) INFORMATION FOR SEQ ID NO:77:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GTTGTTGTTG TTGTTGTTGT T 21

(2) INFORMATION FOR SEQ ID NO:78:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

TAATAATAAT AATAATAATA A 21

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

TACTACTACT ACTACTACTA C

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

TAGTAGTAGT AGTAGTAGTA G 21

(2) INFORMATION FOR SEQ ID NO:81:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:81:

TATTATTATT ATTATTATTA T 21

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:82:

TCATCATCAT CATCATCATC A 21

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

TCCTCCTCCT CCTCCTCCTC C

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

TCGTCGTCGT CGTCGTCGTC G 21

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

TCTTCTTCTT CTTCTTCTTC T 21

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

TGATGATGAT GATGATGATG A 21

(2) INFORMATION FOR SEQ ID NO:87:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

TGCTGCTGCT GCTGCTGCTG C

(2) INFORMATION FOR SEQ ID NO:88:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

TGGTGGTGGT GGTGGTGGTG G 21

(2) INFORMATION FOR SEQ ID NO:89:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

TGTTGTTGTT GTTGTTGTTG T

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

TTATTATTAT TATTATTATT A

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

TTCTTCTTCT TCTTCTTCTT C 21

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

TTGTTGTTGT TGTTGTTGTT G

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

CTCTCTCTCT CTCTCTCTCT 20

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GCAGCAGCAG CAGCAGCAGC AGCA 24

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other nucleic acid

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

GCGGCGGCGG CG 12

WHAT IS CLAIMED IS:

1. Apparatus for identifying tandem nucleotide repeats in a sample containing same, comprising:

a solid support; and

a plurality of oligonucleotides fixedly attached thereto to form an array, the oligonucleotides defining a set of tandem nucleotide repeats including sequences complementary to tandem nucleotide repeats in the sample, the array establishing a pattern such that identity of a tandem nucleotide repeat in the sample may be ascertained by location in the pattern of the complementary tandem nucleotide repeat upon hybridization of the sample to the array.

2. Apparatus according to claim 1, wherein the array comprises a linear sequence of oligonucleotides.

3. Apparatus according to claim 1, wherein the array comprises a two-dimensional pattern.

4. Apparatus according to claim 1, wherein the oligonucleotides have the formula $(Nmer)_n$, in which N is an integer greater than 1 and represents the number of nucleotides in a repeat pattern and n is an integer from two to about 50.

5. Apparatus according to claim 4, wherein the set of tandem nucleotide repeats comprises a complete set of $(3mer)_n$ repeats excluding homopolymers.

6. Apparatus according to claim 1, wherein the set of tandem nucleotide repeats comprises a complete set of $(2mer)_n$ repeats excluding homopolymers.

7. Apparatus according to claim 1, wherein the set of tandem nucleotide repeats comprises a set of oligonucleotides of the formula

$$[(Nmer)_n(Mmer)_m]_x$$

or

$$[(Nmer)_n(Mmer)_m(Pmer)_p]_x$$

in which each of N, M and P is independently selected from the set of integers greater than 1 and represent the number of nucleotides in a repeat pattern;

each of n, m and p is independently selected from the set of integers; and

x is an integer, with the proviso that x(Nn + Mm) or x(Nn + Mm + Pp) is between 4 and about 100.

8. Apparatus according to claim 1, wherein the solid support comprises a material selected from glass and plastic films.

9. Apparatus according to claim 8, wherein the solid support comprises polypropylene.

10. A method for identifying tandem nucleotide repeats in a sample containing same, comprising:

bringing the sample into contact under hybridization conditions with an array comprising a plurality of oligonucleotides fixedly attached to a solid support, the oligonucleotides defining a set of tandem nucleotide repeats including sequences complementary to tandem nucleotide repeats in the sample, the array establishing a predetermined pattern; and

identifying the tandem nucleotide repeats in the sample by determining location in the pattern of the complementary tandem nucleotide repeats to which the tandem nucleotide repeats are hybridized.

11. A method according to claim 10, wherein the sample is labelled prior to bringing it into contact with the array.

12. A method according to claim 11, wherein the label is selected from the group consisting of radioactive labels and fluorescent labels.

13. A method according to claim 10, wherein the oligonucleotides have the formula $(Nmer)_n$, in which N is an integer greater than 1 and represents the number of nucleotides in a repeat pattern and n is an integer from two to about 50.

14. A method according to claim 10, wherein the set of tandem nucleotide repeats comprises a complete set of $(3mer)_n$ repeats excluding homopolymers.

15. A method according to claim 10, wherein the set of tandem nucleotide repeats comprises a complete set of $(2mer)_n$ repeats excluding homopolymers.

16. A method according to claim 10, wherein the set of tandem nucleotide repeats comprises a set of oligonucleotides of the formula

$$[(Nmer)_n(Mmer)_m)]_x$$

or

$$[(Nmer)_n(Mmer)_m(Pmer)_p]_x$$

in which each of N, M and P is independently selected from the set of integers greater than 1 and represent the number of nucleotides in a repeat pattern;

each of n, m and p is independently selected from the set of integers; and

x is an integer, with the proviso that x(Nn + Mm) or x(Nn + Mm + Pp) is between 4 and about 100.

17. A method for determining size of tandem nucleotide repeats in a sample containing same, comprising:

bringing the sample into contact under hybridization conditions with an array comprising a plurality of oligonucleotides fixedly attached to a solid support, the oligonucleotides defining a set of tandem nucleotide repeats including sequences complementary to tandem nucleotide repeats in the sample, the array establishing a predetermined pattern;

identifying the tandem nucleotide repeats in the sample by determining location in the pattern of the complementary tandem nucleotide repeats to which the tandem nucleotide repeats are hybridized to form hybrids;

selectively dissociating the hybrids to elute the tandem nucleotide repeat from the support; and

sizing the eluted tandem nucleotide repeats.

18. A method according to claim 17, wherein the eluted tandem nucleotide repeats are sized by gel electrophoresis against DNA size markers.

	1	2	3	4	10	11–77
1					CYCLE	
2					9	A A A A A A A A A A A A A A A A C C C C C C C C C C C C C C C C G G G G G G G G G G G G G G G G T T T T T T T T T T T T T T T T
3					10	A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T
4					11	A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
5	(ACGT)3-TANDEMREPEA				12	A A A A A A A A A A A A A A A A C C C C C C C C C C C C C C C C G G G G G G G G G G G G G G G G T T T T T T T T T T T T T T T T
6					13	A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T
7					14	A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
8	CYCLE 18				15	A A A A A A A A A A A A A A A A C C C C C C C C C C C C C C C C G G G G G G G G G G G G G G G G T T T T T T T T T T T T T T T T
9					16	A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T
10					17	A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
11					18	A A A A A A A A A A A A A A A A C C C C C C C C C C C C C C C C G G G G G G G G G G G G G G G G T T T T T T T T T T T T T T T T
12					19	A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T A A A A C C C C G G G G T T T T
13					20	A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T A C G T
18	CYCLE →	1	2	3		1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64
19		A	A	A	1	
20		A	A	C	2	
21		A	A	G	3	
22		A	A	T	4	
23		A	C	A	5	
24		A	C	C	6	
25		A	C	G	7	
26		A	C	T	8	
27		A	G	A	9	
28		A	G	C	10	
29		A	G	G	11	
30		A	G	T	12	

FIG. 1(A)

FIG. 1(B)

31	A	T	A	13
32	A	T	C	14
33	A	T	G	15
34	A	T	T	16
35	C	A	A	17
36	C	A	C	18
37	C	A	G	19
38	C	A	T	20
39	C	C	A	21
40	C	C	C	22
41	C	C	G	23
42	C	C	T	24
43	C	G	A	25
44	C	G	C	26
45	C	G	G	27
46	C	G	T	28
47	C	T	A	29
48	C	T	C	30
49	C	T	G	31
50	C	T	T	32
51	G	A	A	33
52	G	A	C	34
53	G	A	G	35
54	G	A	T	36
55	G	C	A	37
56	G	C	C	38
57	G	C	G	39

58	G	C	T	40
59	G	G	A	41
60	G	G	C	42
61	G	G	G	43
62	G	G	T	44
63	G	T	A	45
64	G	T	C	46
65	G	T	G	47
66	G	T	T	48
67	T	A	A	49
68	T	A	C	50
69	T	A	G	51
70	T	A	T	52
71	T	C	A	53
72	T	C	C	54
73	T	C	G	55
74	T	C	T	56
75	T	G	A	57
76	T	G	C	58
77	T	G	G	59
78	T	G	T	60
79	T	T	A	61
80	T	T	C	62
81	T	T	G	63
82	T	T	T	64

FIG. 1(C)

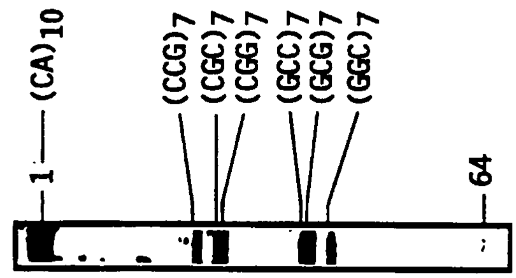


FIG. 2C

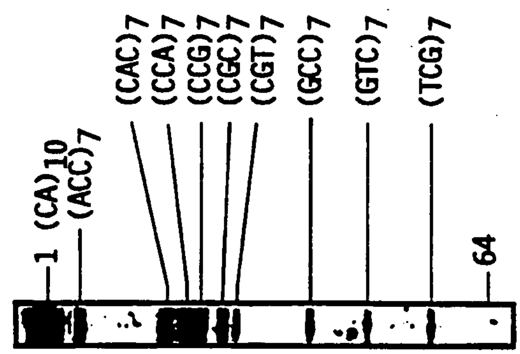


FIG. 2B

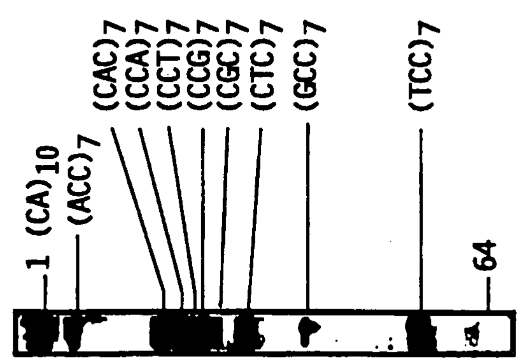
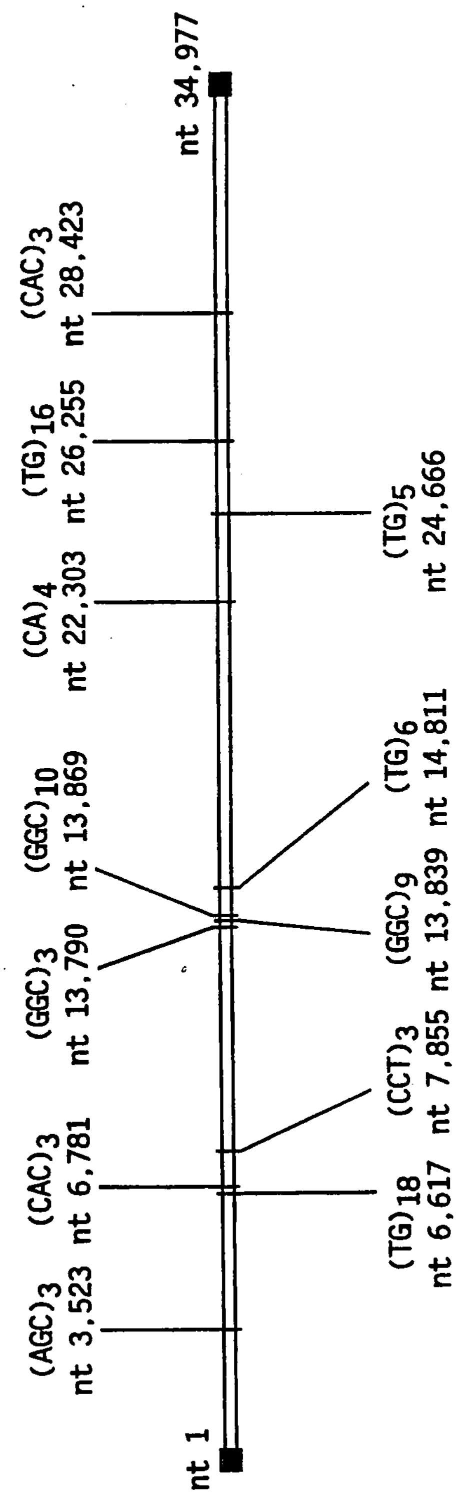


FIG. 2A

FIG. 3


nt 1
(AGC)3
nt 3,523
(CAC)3
nt 6,781
(TG)18
nt 6,617
(CCT)3
nt 7,855
(GGC)3
nt 13,790
(GGC)9
nt 13,839
(GGC)10
nt 13,869
(TG)6
nt 14,811
(CA)4
nt 22,303
(TG)5
nt 24,666
(TG)16
nt 26,255
(CAC)3
nt 28,423
nt 34,977

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12Q1/68

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	WO,A,92 10588 (AFFYMAX TECH NV) 25 June 1992 see the whole document ---	1-18
Y	WO,A,89 10977 (ISIS INNOVATION) 16 November 1989 cited in the application see the whole document ---	1-18
Y	WO,A,93 17126 (PUBLIC HEALTH RESEARCH INST OF) 2 September 1993 see the whole document ---	1-18
	-/--	

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

* Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

1

Date of the actual completion of the international search	Date of mailing of the international search report
4 September 1995	19.09.95

Name and mailing address of the ISA
European Patent Office, P.B. 5818 Patentlaan 2
NL - 2280 HV Rijswijk
Tel. (+31-70) 340-2040, Tx. 31 651 epo nl,
Fax (+31-70) 340-3016

Authorized officer
Molina Galan, E

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	HUMAN MOLECULAR GENETICS, vol. 3, no. 4, April 1994 OXFORD GB, pages 599-605, ARMOUR ET AL. 'Isolation of human simple repeat loci by hybridisation selection' see the whole document ---	1-18
Y	NUCLEIC ACIDS RESEARCH, vol. 21, no. 16, 1993 OXFORD GB, pages 3911-3912, KARAAGYOZOV ET AL. 'Construction of random small-insert genomic libraries highly enriched for simple sequence repeats' see the whole document ---	1-18
A	GENETIC ANALYSIS TECHNIQUES AND APPLICATIONS, vol. 10, no. 1, 1993 NEW YORK US, pages 2-5, IIZUKA ET AL. 'Selective isolation of highly polymorphic (C-A)n * (D-G)n microsatellites by stringent hybridisation' cited in the application ---	
P,X	NUCLEIC ACIDS RESEARCH, vol. 22, no. 9, 11 May 1994 OXFORD GB, pages 1701-1704, WEHNERT ET AL. 'A rapid scanning strip for tri- and dinucleotide short tandem repeats' see the whole document -----	1-18

1

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO-A-9210588	25-06-92	AU-A- 1248292	08-07-92
		EP-A- 0562047	29-09-93
WO-A-8910977	16-11-89	AT-T- 110790	15-09-94
		DE-D- 68917879	06-10-94
		DE-T- 68917879	05-01-95
		EP-A- 0373203	20-06-90
		JP-T- 3505157	14-11-91
WO-A-9317126	02-09-93	AU-B- 3728093	13-09-93
		CA-A- 2130562	02-09-93